Codes, orderings, and partial words

By: F. Blanchet-Sadri

F. Blanchet-Sadri, "Codes, Orderings, and Partial Words." *Theoretical Computer Science*, Vol. 329, 2004, pp 177-202. DOI: 10.1016/j.tcs.2004.08.011

Made available courtesy of Elsevier: http://dx.doi.org/10.1016/j.tcs.2004.08.011

Abstract:
Codes play an important role in the study of the combinatorics of words. In this paper, we introduce pcodes that play a role in the study of combinatorics ofpartial words. Partial words are strings over a finite alphabet that may contain a number of "do not know" symbols. Pcodes are defined in terms of the compatibility relation that considers two strings over the same alphabet that are equal except for a number of insertions and/or deletions of symbols. We describe various ways of defining and analyzing pcodes. In particular, many pcodes can be obtained as antichains with respect to certain partial orderings. Using a technique related to dominoes, we show that the pcode property is decidable.

Keywords: Word; Partial word; Partial ordering; Code; Antichain; Domino

Article:
1. Introduction
The theory of codes has been widely developed in connection with combinatorics on words [2]. In this paper, we introduce *pcodes* in connection with combinatorics on *partial words*. Pcodes are defined in terms of the *compatibility relation* which considers two sequences over the same alphabet that are equal except for a number of insertions and/or deletions. We describe various ways of defining and analyzing pcodes. In particular, many pcodes can be obtained as antichains with respect to some special partial orderings. We show that the pcode property can be decided for finite sets of partial words. The decidability result for pcodes is an adaptation of the domino graph technique of Head and Weber [13].

A motivation for considering partial words comes from the study of biological sequences such as DNA and protein that play a central role in molecular biology. DNA sequences can be viewed as long (a few million to a few billion letters) strings in the 4-letter alphabet of *nucleotides*: *a* (for adenine), *c* (for cytosine), *g* (for guanine), and *t* (for thymine), while protein sequences can be viewed as short (a few hundred letters) strings in the 20-letter alphabet of *amino acids*. Proteins are made by fragments of DNA called *genes* that are roughly three times longer than the corresponding proteins. This is because every triplet of nucleotides in the DNA alphabet codes one letter in the protein alphabet of amino acids.

Sequence comparison is one of the most important primitive operation in molecular biology, serving as a basis for many other, more complex, manipulations. Alignment of two sequences is a way of placing one sequence above the other in order to make clear the correspondence between similar letters or substrings from the sequences. Alignment of two genes (or two proteins) can be viewed as a construction of two partial words that are said to be compatible. As an example, consider the sequences *gacggattag* and *gatcggtag*. We cannot help but notice that they actually look very much alike, a fact that becomes more obvious when we align them one above the other as follows:

$$\begin{array}{cccccccccc} g & a & \diamond & c & g & g & a & t & t & a & g \\ g & a & t & c & g & g & \diamond & \diamond & t & a & g \end{array}$$

The second sequence is obtained from the first by inserting a *t* and by deleting an *a* and a *t*. Observe that we had to introduce gaps or holes (indicated by ◊'s) in the sequences to let similar nucleotides align perfectly.

Another important operation in molecular biology where partial words play a role is DNA sequencing. DNA sequencing is the process of obtaining from a DNA molecule its base sequence. The computational task involved in DNA sequencing is called fragment assembly of DNA. The motivation for this problem comes from the fact that with current technology it is impossible to sequence directly contiguous stretches of more than a few hundred bases. On the other hand, there is technology to cut random pieces of a long DNA molecule and to produce enough copies of the pieces to sequence. Thus, a typical approach to sequencing long DNA molecules is to sample and then sequence fragments from them. However, this leaves us with the problem of assembling the pieces. As an example, suppose the input is composed of the four sequences *accgt*, *cgtgc*, *ttac*, *taccgt* and we know that the answer has approximately 10 bases. One possible way to assemble this set is

$$\begin{array}{cccccccccc} \diamond & \diamond & a & c & c & g & t & \diamond & \diamond \\ \diamond & \diamond & \diamond & \diamond & c & g & t & g & c \\ t & t & a & c & \diamond & \diamond & \diamond & \diamond & \diamond \\ \diamond & t & a & c & c & g & t & \diamond & \diamond \end{array}$$

which gives us *ttaccgtgc*. This answer has 9 bases, which is close to the target length of 10. The only guidance to assembly, apart from the approximate size of the target, are the overlaps between fragments. By overlap we mean the fact that sometimes the end part of a fragment is similar to the beginning of another, as with the first and second sequences above. Again we had to introduce gaps or holes (indicated by ◊'s) in the sequences to let similar bases before and after the ◊'s align perfectly. Real problem instances however are very large. Apart from this fact, several other complications exist that make the problem much harder than the small example above. The main factors that add to the complexity of the problem are errors, regularities such as periodicities and repetitions, and lack of coverage [17]. Research in combinatorics of partial words was initiated by Berstel and Boasson [1]. Other works include [4–9].

This paper studies codes, orderings, and partial words. In Section 2, notation and basic notions on words and partial words are discussed. In particular, the roles of compatibility and commutativity are investigated. In Section 3, some special basic binary relations are defined on partial words including the prefix, suffix, commutative, and border relations. There, the role of primitivity of partial words is also discussed. In Section 4, pcodes are introduced and their properties concerning binary relations are proved. In Section 5, the class of antichains with respect to the prefix and suffix partial orderings of partial words is characterized. In Section 6, the border partial ordering is discussed. Section 7 contains results related to the commutative partial ordering on partial words. Moreover, in Section 8 we show that the pcode property is decidable. Finally, Section 9 contains a few concluding remarks.

2. Preliminaries

In this section, we recall some basic notions on words and partial words.

Let A be a nonempty finite set of symbols called an *alphabet*. Symbols in A are called *letters* and any finite string over A is called a *word* over A. The *empty word*, that is the word containing no letter, is denoted by ε. The set of all words over A is denoted by A^*. It is a monoid under the associative operation of concatenation or product of words, and ε serves as identity. We call $A^+ = A^* \setminus \{\varepsilon\}$ the *free semigroup* generated by A and A^* the *free monoid* generated by A.

A word of length n over A can be defined by a total function $u: \{0,\ldots, n-1\} \to A$ and is usually represented as $u = a_0a_1\ldots a_{n-1}$ with $a_i \in A$. A *partial word* of length n over A is a partial function $u: \{0,\ldots, n-1\} \to A$. For $0 \le i < n$, if $u(i)$ is defined, then we say that i belongs to the *domain* of u (denoted by $i \in D(u)$), otherwise we say that i belongs to the *set of holes* of u (denoted by $i \in H(u)$). A word over A is a partial word over A with an empty set of holes (we sometimes refer to words as *full* words). For any partial word u over A, $|u|$ denotes its length. In particular, $|\varepsilon| = 0$. We denote by W_0 the set A^*, and for every integer $i \geq 1$, by W_i the set of partial words over A with at most i holes. We put $W = \cup_{i\geq 0} Wi$, the set of all partial words over A with an arbitrary number of holes.

If u is a partial word of length n over A, then the *companion* of u (denoted by $u_\Diamond$) is the total function $u_\Diamond : \{0,..., n-1\} \to A \cup \{\Diamond\}$ defined by

$$u_\Diamond(i) = \begin{cases} u\,(i) if\ i\ \in\ D\,(u), \\ \Diamond\ otherwise \end{cases}$$

The symbol $\Diamond \notin A$ is viewed as a “do not know” symbol. The word $u_\Diamond = abb\Diamond b\Diamond cb$ is the companion of the partial word u of length 8 where $D(u) = \{0, 1, 2, 4, 6, 7\}$ and $H(u) = \{3, 5\}$. The bijectivity of the map $u \mapsto u_\Diamond$ allows us to define for partial words concepts such as concatenation and powers in a trivial way. The set W is a monoid under the concatenation (ε serves as identity). For a word u, the powers of u are defined inductively by $u^0 = \varepsilon$ and, for any $n \geq 1$, $u^n = uu^{n-1}$. For a subset X of W, we denote by X^* the submonoid of W generated by X. It consists of all partial words which are concatenations of elements of X.

A partial word u is a *factor* of the partial word v if there exist partial words x, y such that $v = xuy$. The factor u is called *proper* if $u \neq v$. The partial word u is a *prefix* (respectively, *suffix*) of v if $x = \varepsilon$ (respectively, $y = \varepsilon$). For a subset X of W, we denote by $F(X)$ the set of factors of elements in X. More specifically,

$F(X) = \{u \mid u \in W$ and there exist $x, y \in W$ such that $xuy \in X\}$.

A *period* of a partial word u over A is a positive integer p such that $u\,(i) = u\,(j)$ whenever $i, j \in D\,(u)$ and $i \equiv j$ mod p. In such a case, we call u *p-periodic*.

For convenience in the sequel, we consider a partial word over A as a word over the enlarged alphabet $A \cup \{\Diamond\}$, where the additional symbol $\Diamond$ plays a special role. Thus, we say for instance “the partial word $\Diamond ab\Diamond b$” instead of “the partial word with companion $\Diamond ab\Diamond b$”.

2. 1. Compatibility

In this section, we discuss compatibility on partial words.

If u and v are two partial words of equal length, then u is said to be contained in v, denoted by $u \subset v$, if all elements in $D(u)$ are in $D(v)$ and $u(i) = v(i)$ for all $i \in D(u)$. We sometimes write $u \sqsubset v$ if $u \subset v$ but $u \neq v$. The order $u \subset v$ on partial words is obtained when we let $\Diamond \prec a$ and $a \preceq a$ for all $a \in A$. The partial words u and v are called *compatible*, denoted by $u \uparrow v$, if there exists a partial word w such that $u \subset w$ and $v \subset w$. We denote by $u \vee v$ the least upper bound of u and v (in other words, $u \subset u \vee v$ and $v \subset u \vee v$ and $D(u \vee v) = D(u) \cup D(v)$). As an example, $u = aba\Diamond\Diamond a$ and $v = a\Diamond\Diamond b\Diamond a$ are two partial words that are compatible and $u \vee v = abab\Diamond a$. For a subset X of W, we denote by $C(X)$ the set of all partial words compatible with elements of X. More specifically,

$C(X) = \{u \mid u \in W$ and there exists $v \in X$ such that $u \uparrow v\}$.

The following rules are useful for computing with partial words.
Multiplication: If $u \uparrow v$ and $x \uparrow y$, then $ux \uparrow vy$.
Simplification: If $ux \uparrow vy$ and $|u| = |v|$, then $u \uparrow v$ and $x \uparrow y$.
Weakening: If $u \uparrow v$ and $w \subset u$, then $w \uparrow v$.

Lemma 1 (*Berstel and Boasson* [1]). *Let $u, v, x, y \in W$ be such that $ux \uparrow vy$.*

- *If $|u| > |v|$, then there exist $w, z \in W$ such that $u = wz$, $v \uparrow w$, and $y \uparrow zx$.*

- *If* $|u| <, |v|$, *then there exist* $w, z \in W$ *such that* $v = wz$, $u \uparrow w$, *and* $x \uparrow zy$.

2.2. Commutativity
In this section, we discuss commutativity on partial words.

Lemma 2 (*Shyr [18]*). *Let* $u, v \in W_0 \backslash \{\varepsilon\}$. *If* $uv = vu$, *then there exists* $w \in W_0$ *such that* $u = w^m$ *and* $v = w^n$ *for some integers* m, n.

We now describe an extension of Lemma 2.

Definition 1 (*Blanchet-Sadri and Luhmann [9]*). Let k, be positive integers satisfying $k \leq l$ For $0 \leq i < k + l$ we define the sequence of i relative to k, as $seq_{k,l}(i) = (i_0, i_1, i_2, \ldots, i_n, i_{n+1})$ where

- $n = ((k + l)/gcd(k,l)) - 1$,
- $i_0 = i = i_{n+1}$,
- For $1 \leq j \leq n$, $ij \neq i$,
- For $1 \leq j \leq n + 1$, i_j is defined as

$$ij = \begin{cases} i_{j-1} + k \text{ if } i_{j-1} < l, \\ i_{j-1} - l \text{ otherwise} \end{cases}$$

For example, if $k = 4$ and $l = 10$, then $seq_{4,10}(0) = (0, 4, 8, 12, 2, 6, 10, 0)$.
Definition 2 (*Blanchet-Sadri and Luhmann [9]*). Let k, be positive integers satisfying $k \leq l$, and let $w \in W$ be of length $k + l$. We say that w is $\{k, l\}$-special if there exists $0 \leq i < k$ such that $seq_{k,l}(i) = (i_0, i_1, i_2, \ldots, i_n, i_{n+1})$ satisfies one of the following conditions:

- $seqk,l\ (i)$ contains two consecutive positions that are holes of w.
- $seqk,l\ (i)$ contains two positions that are holes of w while $w_\Diamond(i_0)w_\Diamond(i_1)w_\Diamond(i_2)\ldots w_\Diamond(i_{n+1})$ is not 1-periodic.

For example, if $k = 4$ and $l = 10$, then

- The partial word $u = ab\Diamond aab\Diamond aabaa\Diamond\Diamond$ is $\{4, 10\}$-special since $seq_{4,10}(0)$ contains the consecutive positions 12 and 2 which are in $H(u) = \{2, 6, 12, 13\}$.
- The partial word $v = a\Diamond baab\Diamond aabaa\Diamond\Diamond$ is $\{4, 10\}$-special since $seq_{4,10}(0)$ contains the positions 6 and 12 which are in $H(v) = \{1, 6, 12, 13\}$ while

$$v_\Diamond(0)v_\Diamond(4)v_\Diamond(8)v_\Diamond(12)v_\Diamond(2)v_\Diamond(6)v_\Diamond(10)v_\Diamond(0) = aaa\Diamond b\Diamond aa$$

is not 1-periodic.

- The partial word $w = \Diamond babab\Diamond babab\Diamond b$ is not $\{4, 10\}$-special.

The following lemmas were used to prove Theorem 1 that follows.

Lemma 3 (*Blanchet-Sadri and Luhmann [9]*). *Let* $x, y \in W_0 \backslash \{\varepsilon\}$, *and let* $w \in W$ *be non* $\{|x|, |y|\}$*-special. If* $w \subset xy$ *and* $w \subset yx$, *then* $xy = yx$.

The special case of Lemma 3 when w has only one hole was proved in [1]. In this case, w is by definition non $\{|x|, |y|\}$-special.

Lemma 4 (*Blanchet-Sadri and Luhmann [9]*). *Let* $u, v \in W \backslash \{\varepsilon\}$, $x, y \in W_0$ *be such that* $u \subset x$ *and* $v \subset y$. *If* uv *is non* $\{|u|, |v|\}$*-special and* $vu \subset xy$, *then* $uv \subset yx$.

Theorem 1 (*Blanchet-Sadri and Luhmann* [9]). *Let $u, v \in W \backslash \{\varepsilon\}$ be such that uv is non* $\{|u|, |v|\}$*-special. If* $uv \uparrow vu$*, then there exists* $w \in W_0$ *such that* $u \subset w^m$ *and* $v \subset w^n$ *for some integers m, n.*

The special case of Theorem 1 when uv has only one hole was proved in [1]. Theorem 1 does not hold if uv is $\{|u|, |v|\}$-special. Take for example $u = \Diamond bb$ and $v = abb\Diamond$. We have $uv \uparrow vu$ but no word w and no integers m, n satisfying $u \subset w^m$ and $v \subset w^m$. Note that $seq_{3,4}(0) = (0, 3, 6, 2, 5, 1, 4, 0)$ contains the holes 0, 6 of uv while

$$(uv),(0)(uv),(3)(uv),(6)(uv),(2)(uv),(5)(uv),(1)(uv),(4)(uv),(0)$$
$$= \Diamond a \Diamond bbbb\Diamond$$

is not 1-periodic showing that $u\,v$ is j3, 4}-special.

We end this section with the concept of a *pairwise nonspecial set of partial words* that is used in the sequel.

Definition 3. Let $X \subset W$. Then X is called *pairwise nonspecial* if all $u, v \in X$ of different positive lengths satisfy the following conditions:

- If $|u| < |v|$, then v is non $\{|u|, |v| - |u|\}$-special.
- If $|u| > |v|$, then u is non $\{|v|, |u| - |v|\}$-special.

Note that any subset of W_1 is pairwise nonspecial.

3. Binary relations

Throughout, we fix a finite alphabet A. We assume that the cardinality of A, denoted by $||A||$, is at least two (unless stated otherwise).

A binary relation p defined on an arbitrary set $S \subset W$ is a subset of $S \times S$. Instead of denoting $(u, v) \in p$, we often write upv. The relation p is called *reflexive* if upu for all $u \in S$; *symmetric* if upv implies vpu for all $u, v \in S$; *antisymmetric* if upv and vpu imply $u = v$ for all $u, v \in S$; *transitive* if upv and vpw imply upw for all $u, v, w \in S$, and *positive* if $\varepsilon\ \varepsilon pu$ for all $u \in S$. It is called *strict* if it satisfies the following conditions for all $u, v \in S$:
upu,
upv implies $|u| \leq |v|$,
upv and $|u| = |v|$ imply $v \subset u$.

A strict binary relation is reflexive and antisymmetric, but not necessarily transitive. A reflexive, antisymmetric, and transitive relation p defined on S is called a *partial ordering*, and (S, p) is called a *partially ordered set*. A partial ordering p on S is called *right* (respectively, *left*) *compatible* if upv implies $uwpvw$ (respectively, upv implies $wupwv$) for all $u, v, w \in S$. It is called *compatible* if it is both right and left compatible. For any two binary relations p_1 and p_2 on S, we denote by $(p_1)\ (p_2)$ if $up1\ v$ implies up_2v for all $u, v \in S$ (or the subset inclusion), and by $(p_1) \sqsubset (p_2)$ if $(p_1) \subset (p_2)$ but $(p_1) \neq (p_2)$.

An important notion on binary relations is that of an *antichain*. A nonempty subset X of S is called an *antichain* with respect to a particular binary relation p on S (or an p-antichain) if for all distinct $u, v \in X$, $(u, v) \notin p$ and $(v, u) \notin p$. The class of all p-antichains of S is denoted by $A(p)$. For every partial word u of S, $\{u\}$ is in $A(p)$.

Proposition 1. *Let p_1, p_2 be two binary relations defined on W. Then*

1. *If $p_1 \subset p_2$, then $A(p_2) \subset A(p_1)$.*
2. *If p_1, p_2 are strict and $A(p_2) \subset A(p_1)$, then $p_1 \subset p_2$.*

Proof. For Statement 1, let $X \in A(p_2)$. If X is a singleton set, then $X \in A(p_1)$. Now suppose that X is not a singleton set and let $u, v \in X$ be such that $u \neq v$ and up_1v. Then up_2v by assumption. Since X is an antichain with respect to p_2, we have $u = v$, a contradiction. Thus $X \in A(p_1)$ and $A(p_2) \subset A(p_1)$ holds.

For Statement2, suppose that there exist $u, v \in W$ such that $u \neq v$, up_1v, and $(u, v) \notin p2$. Suppose that vp_2u. Since up_1v, we have $|u| \leq |v|$, and since vp_2u, we have $|v| \leq |u|$. Hence $|u| = |v|$, both $up_1 v$ and $|u| = |v|$ imply $v \subset u$, and both vp_2u and $|v| = |u|$ imply $u \subset v$. We deduce that $u = v$, a contradiction. So $\{u, v\} \in A(p2)$. As $A(p2) \subset A(p1)$, we have $\{u, v\} \in A(p\ 1)$ which implies that $(u, v) \neq p_1$, a contradiction. ■

3.1. The δ-relations

A word $u \in W_0\backslash\{\varepsilon\}$ is *primitive* if $u = v^n$ for some $v \in W_0\backslash\{\varepsilon\}$ implies $n = 1$. Note the fact that the empty word ε is not primitive. For $u \in W_0\backslash\{\varepsilon\}$, there exists a unique primitive word $v \in W_0\backslash\{\varepsilon\}$ and a unique positive integer n such that $u = v^n$. We call v *the* (*primitive*) *root* of u, and denote it by $\sqrt{u}$. All positive powers of u have the same root. For $u, v \in W_0\backslash\{\varepsilon\}$, $uv = vu$ is equivalent to $\sqrt{u} = \sqrt{v}$. For more details on these results, we refer the reader to [18] for instance.

A partial word $u \in W\backslash\{\varepsilon\}$ is *primitive* if $u \subset v^n$ for some $v \in W_0\backslash\{\varepsilon\}$ implies $n = 1$. Note that if x is primitive and $x \subset y$, then y is primitive as well. For $u \in W\backslash\{\varepsilon\}$, there exists a primitive word $v \in W_0\backslash\{\varepsilon\}$ and a positive integer n such that $u \subset v^n$. However, uniqueness does not hold as is seen with the partial word $u = a\Diamond$ ($u \subset a^2$, $u \subset ab$ with distinct letters a, b and both a, ab are primitive). For $u \in W\backslash\{\varepsilon\}$, let $P(u)$ denote the set of primitive words $v \in W_0\backslash\{\varepsilon\}$ such that $u \subset v^n$ for some positive integer n. For $u \in W_0\backslash\{\varepsilon\}$, we have $P(u^i) = P(u) = \{\sqrt{u}\}$, and for $u \in W\backslash\{\varepsilon\}$, we have $P(u) \subset P(u^i)$ for all positive powers of u.

For every positive integers i, j and partial words $u, v \in W\backslash\{\varepsilon\}$, define the relation $\delta_{i,j}$ by $u\delta_{i,j}v$ if $P(u^i) \cap P(v^j) \neq \emptyset$. In the sequel, $\delta_{1,1}$ is often abbreviated by δ. Note that if $u \subset v$, then $P(v) \subset P(u)$ and so ubv.

Lemma 5. *Let i, j be positive integers.*

1. *If* $||A|| \geq 2$, *then* $(\delta) \subset (\delta_{i,j})$. *Moreover, if* $(i, j) \neq (1, 1)$, *then* $(\delta) \sqsubset (\delta_{i,j})$.
2. *If* $||A|| = 1$, *then* $(b) = (bi,j)$.

Proof. The inclusion $\subset$ in Statement 1 follows from the fact that $P(u) \subset P(u^k)$ for all u, k. To see that the inclusion $\sqsubset$ holds in case $(i, j) \neq (1, 1)$, we argue as follows: if $i > 1$, then we consider $u = \Diamond$ and $v = a^{i-1}b$ which satisfy $(u , v) \notin \delta$ and $u\delta_{i,j}v$; and if $j > 1$, then we consider $u = a^{j-1}b$ and $v = \Diamond$ which satisfy $(u, v) \notin \delta$ and $u\delta_{i,j}v$. Statement 2 follows from the fact that $P(u) = A$ for all u. ■

Lemma 6. *Let i, j be positive integers, and let* $u, v \in W\backslash\{\varepsilon\}$.

1. *If* $u\delta_{i,j}v$, *then* $u^iv^j \uparrow v^ju^i$.
2. *If* $u^iv^j \uparrow v^ju^i$ *and* u^iv^j *is non* $\{|ui|, |vj|\}$*-special, then* $u\delta_{i,j}v$.

Proof. For Statement 1, if $u\delta_{i,j}v$, then let $w \in P(u^i) \cap P(v^j)$. Then $u^i \subset w^m$ and $v^j \subset w^n$ for some integers m, n. We have $u^iv^j \subset w^{m+n}$ and $v^ju^i \subset w^{m+n}$, and $u^iv^j \uparrow v^ju^i$ follows. For Statement 2, by definition there exists a word w such that $u^iv^j \subset w$ and $v^j u^i \subset w$. Put $w = xy$ where $|x| = |u^i|$ and $|y| = |v^j|$. Since u^iv^j is non $\{|u^i|, |v^j|\}$-special, we get $u^i v^j \subset y^x$ by Lemma 4. The two inclusions $u^i v^j \subset xy$, $u^iv^j \subset yx$ give $xy = yx$ by Lemma 3, and thus $\sqrt{x} = \sqrt{y}$. Hence $\sqrt{x} \in P(u^i) \cap P(v^j)$ since $u^i \subset x$ and $v^j \subset y$. The relation $u\delta_{i,j}v$ follows. ■

3.2. The p-relations

The following are some useful binary relations on W that generalize some well-known binary relations on W_0.

Definition 4. Let $u, v \in W$.

- *Embedding relation*: up_dv if there exists an integer $n \geq 0$, $u_1,..., u_n \in W$,and $x_0,...,x_n \in W_0$ such that $u = u_1u_2...u_n$ and $v \subset x_0u_1x_1u_2...u_nx_n$.
- *Length relation*: up_lv if $|u| < |v|$ or $v \subset u$.
- *Prefix relation*: up_pv if there exists $x \in W_0$ such that $v \subset ux$.

- *Suffix relation*: $u p_s v$ if there exists $x \in W_0$ such that $v \subset xu$.
- *Factor relation*: $u p_f v$ if there exist $x, y \in W_0$ such that $v \subset xuy$.
- *Border relation*: $u p_o v$ if there exist $x, y \in W_0$ such that $v \subset ux$ and $v \subset yu$.
- *Commutative relation*: $u p_c v$ if there exists $x \in W_0$ such that $v \subset xu$, $v \subset ux$.
- *Exponent relation*: $u p_e v$ if there exists an integer $n \geq 1$ such that $v \subset u^n$.

Lemma 7.
- *The relations p_d, p_l, p_p, p_s, p_f, and p_o are strictpositive partial orderings on W.*
- *The relation p_e is a strictpartial ordering on W.*
- *The relation p_c is a strictpositive binary relation on W.*
- *The relation p_c is a partial ordering on anypairwise nonspecial subset of W.*

Proof. We show the result for the relation p_c (the proofs for the other relations are straight- forward). The relation p_c is trivially strict and positive on W. Now, let X be a pairwise nonspecial subset of W. To show that p_c is transitive, let $u, v, w \in X$ be such that $u \neq v$ and $v \neq w$. If $u p_c v$ and $v p_c w$, then let us show that $u p_c w$. If $u = \varepsilon$, then trivially $\varepsilon p_c w$, and if $v = \varepsilon$, then $u = \varepsilon$. So we assume that u, v are nonempty. For some words x and y, we have $v \subset xu$, $v \subset ux$ and $w \subset vy$, $w \subset yv$. If $x = \varepsilon$, then $v \subset u$. We get $w \subset vy \subset uy$ and $w \subset yv \subset yu$, and so $u p_c w$. If $y = \varepsilon$, then $w \subset v$. We get $w \subset xu$ and $w \subset ux$, and so $u p_c w$. So we may assume that x, y are nonempty. Let u' be a full word satisfying $u \subset u'$. We get $v \subset xu'$, $v \subset u'x$ and thus by Lemma 3, $xu' = u'x$. By Lemma 2, there exists a primitive word z (we can choose $z = \sqrt{x}$) and positive integers k,l such that $u' = z^k$ and $x = z^l$. We have $v \subset u'x \subset z^{k+l}$. We get $w \subset z^{k+l} y$, $w \subset yz^{k+l}$. Thus by Lemma 3, $z^{k+l} y = yz^{k+l}$. Using Lemma 2 and the fact that z is primitive, we get that y is a power of z, say $y = z^m$ for some integer m. It follows that $w \subset vy \subset uxy \subset uz^{l+m}$ and also $w \subset yv \subset yxu \subset z^{l+m}u$, and so $u p_c w$. ■

Lemma 8.
- *If* $\|A\| \geq 2$, *then* $(p_c) \sqsubset (p_o)$.
- *If* $\|A\| \geq 1$, *then*
  $(p_e) \sqsubset (p_c) \subset (p_o) \sqsubset (p_p) \sqsubset (p_f) \sqsubset (p_d) \sqsubset (p_l)$, *and*
  $(p_o) \sqsubset (p_s) \sqsubset (p_f)$.
- *If* $\|A\| = 1$, *then* $(p_c) = (p_o)$.

Proof. If $A = \{a\}$, then $(u, v) \notin p_e$ and $u p_c v$ with $u = aa$ and $v = a\Diamond a$, $(u, v) \notin p_o$ and $u p_p v$ with $u = \Diamond$ and $v = \Diamond aa$, $(u, v) \notin p_p$ and $u p_f v$ with $u = \Diamond$ and $v = a\Diamond a$, $(u, v) \notin p_f$ and $u p_d v$ with $u = \Diamond\Diamond$ and $v = \Diamond a\Diamond$, $(u, v) \notin p_d$ and $u p_l v$ with $u = \Diamond$ and $v = aa$, $(u, v) \notin p_o$ and $u p_s v$ with $u = \Diamond$ and $v = a\Diamond$, and $(u, v) \notin p_s$ and $u p_f v$ with $u = \Diamond$ and $v = a\Diamond a$. ■

Note that if we restrict ourselves to $W1$ and $11A11 = 1$, we have $(pf) = (pd)$.

Proposition 2. *The embedding relation p_d is the smallest positive compatible partial ordering on W satisfying $a p_d \Diamond$ for all $a \in A$, that is, if p is a positive compatible partial ordering on Wsatisfying $a p \Diamond$ for all $a \in A$, then* $(p_d) \subset (p)$.

Proof. The embedding partial ordering p_d is clearly compatible on W. Now, let p be a positive compatible partial ordering on $W$and let $u, v \in W$ be such that $u p_d v$. By induction on $|u| + |v|$, we show that upv. If $|u| + |v| = 0$, then $\varepsilon p_d \varepsilon$ and $\varepsilon p \varepsilon$ since p_d and p are positive. If $|u| + |v| > 0$ and $u = \varepsilon$, then $\varepsilon p_d v$ and $\varepsilon p v$ since p_d and p are positive. If $|u| + |v| > 0$ and $u \neq \varepsilon$, then put $u = au'$ and $v = bv'$ where $a, b \in A \subset \{\Diamond\}$. If $a = b$, then $u' p_d v'$, and using the inductive hypothesis, we get $u' p v'$. Since p is compatible, we have $au' p av'$ and so upv. If $a \neq b$ and $b \neq \Diamond$, then $u p_d v'$, and thus by the inductive hypothesis, $u p v'$. Since p is positive, we have $\varepsilon p b$ and since p is compatible, we have $v' p b v'$ and so $v' p v$. Since p is transitive, we get upv as desired. On the other hand, if $a \neq b$ and $b = \Diamond$, then $u' p d v'$, and thus by the inductive hypothesis, $u' p v'$. Since $a p \Diamond$ and p is compatible, we have

$av'p\Diamond v'$. Since $u'pv'$ and p is compatible, we have $au'pav'$. Since p is transitive, we get $au'p\Diamond v'$ or upv as desired. ■

Theoretical aspects of the embedding ordering on W_0 can be found in [10,12,14,16]. An algorithmic aspect of the embedding ordering is motivated by molecular biology. The problem is to find, for a given set $X = \{u_1,...,u_n\}$ of words, a shortest word v such that $u_i p_d v$ for all i. This problem is referred to as the shortest common supersequence problem which is known to be NP-complete [17].

4. Codes

In this section, we extend the notion of code of words to pcode of partial words.

Let X be a nonempty subset of $W0\backslash\{\varepsilon\}$. Then X is called a *code* if for all integers $m \geq 1$, $n \geq 1$ and words $u_1,...,u_m$, $v1,...,v_n \in X$, the condition

$$u_1u_2...u_m = v_1v_2...v_n$$

implies $m = n$ and $u_i = v_i$ for $i = 1,...,m$.

In the case of partial words, we define a *pcode* as follows.

Definition 5. Let X be a nonempty subset of $W\backslash\{\varepsilon\}$. Then X is called a *pcode* if for all integers $m \geq 1$, $n \geq 1$ and partial words $u_1,...,u_m$, $v_1,...,v_n \in X$, the condition

$$u_1u_2...u_m \uparrow v_1v_2...v_n$$

implies $m = n$ and $ui = vi$ for $i = 1,. . . ,m$.

It is clear from the definition that a subset X of $W_0\backslash\{\varepsilon\}$ is a code if and only if it is a pcode. The following proposition extends a property of codes [18].

Proposition 3. *Let X be a nonempty subset of* $W\backslash\{\varepsilon\}$. *Then X is a pcode if and only if for every integer* $n > 1$ *and partial words* $u_1,..., u_n$, v_1 ,...,$v_n \in X$, *the condition*

$$u_1u_2 \; ... \; u_n \uparrow v_1v_2 \; ... \; v_n$$

implies $u_i = v_i$ *for* $i = 1,...,n$.

Proof. If$X$is a pcode, then clearly the condition holds. Conversely, assume that X satisfies the condition stated in the proposition. Suppose $u_1u_2...u_m \uparrow v_1v_2 \; ... \; vn$ for some integers $m \geq 1$, $n \geq 1$ and partial words $u_1,...,u_m$, $v_1,...,v_n \in X$. Then

$$u_1u_2...u_mv_1 \; v_2...v_n \uparrow v_1v_2...v_nu_1u_2...u_m$$

by multiplication. If $m < n$, then $u_1 = v_1$, ... ,$u_m = v_m$ and $\varepsilon \uparrow v_{m+1}...v_n$, which is a contradiction. Similarly, $n < m$ cannot hold. Hence $m = n$ and therefore the condition implies that $X$is a pcode. ■

Proposition 4. *Let X be a nonempty subset of* $W\backslash\{\varepsilon\}$. *For every* $u \in X$, *let* $x_u \in W\backslash\{\varepsilon\}$ *be such that* $u \subset x_u$, *and let Y be the set* $\{x_u \mid u \in X\}$. *If X is a pcode, then Y is a pcode.*

Proof. Let n be a positive integer and let $x_1,...,x_n,y_1,...,y_n \in Y$ be such that

$$x_1x_2 \; ... \; x_n \uparrow y_1y_2 \; ... \; y_n.$$

For every integer $1 \leq i \leq n$, let $u_i \in X$ be such that $x_{u_i} = x_i$, and let $v_i \in X$ be such that $x_{v_i} = y_i$. Then we have

$u_1u_2 \ldots u_n \uparrow v_1v_2 \ldots v_n$

since $u_1u_2 \ldots u_n \subset x_1x_2 \ldots x_n \subset w$ and $v_1v_2 \ldots v_n \subset y_1y_2 \ldots y_n \subset w$ for some w. But since X is a pcode, by Proposition 3, $u_i = v_i$ for $i = 1,\ldots,n$. This implies $x_i = x_{u_i} = x_{v_i} = y_i$ for $i = 1,\ldots,n$ showing that Y is a pcode. ■

The converse of this proposition is not true. For example, let $X = \{u, v\}$ where $u = a$ and $v = a\Diamond a$. The set $Y = \{a, aba\}$ is a pcode, but X is not a pcode since $u^3 \uparrow v$.

The following proposition shows that there is no strict positive binary relation p with the class ofpcodes being the class of p-antichains.

Proposition 5. *If $||A|| \geq 2$, then there is no strict positive binary relation p defined on W such that $A(p)$ is exactly the class of all pcodes over A.*

Proof. Suppose to the contrary that p is a strict positive binary relation defined on W such that $A(p)$ is exactly the class of all pcodes over A. Then the restriction of p to W_0 is a strict positive binary relation such that the class ofall p-antichains is exactly the class of all codes over A contradicting a result of [19]. ■

4.1. The class F

We now consider the following class of binary relations on W partially ordered by inclusion:

$F = \{p \mid p$ is a strict binary relation on W such that every pcode is an antichainwith respect to $p\}$.

The class F is easily seen to be closed under union and intersection. It was considered in [19] for strict positive binary relations on W_0.

The following proposition gives some closure properties for F.

Proposition 6. *Let y be a strict binary relation on W and let $p \in F$. Then the following conditions hold*:

1. *If $(y) \subset (p)$, then $y \in F$,*
2. *$y \cap p \in F$.*

Proof. Statement 1 follows immediately from Proposition 1. For Statement 2, since $(y \cap p) \subset (p)$ and $y \cap p$ is strict, then $y \cap p \in F$ follows from Statement 1. ■

The next proposition implies that $(\delta_{i,j} \cap p) \in F$ for all positive integers i, j and every strict binary relation p on W.

Proposition 7. *Let P be a strict binary relation on W, let X be a nonempty subset of $W \backslash \{\varepsilon\}$, and let i, j be positive integers. If X is a pcode, then X is an $(\delta_{i,j} \cap p)$-antichain.*

Proof. Let X be a pcode. The case where X contains only one partial word is trivial. So let $u, v \in X$ be such that $u \neq v$ and $u\ (\delta_{i,j} \cap P)\ v$. The latter yields $u\delta_{i,j}v$ and by Lemma 6(1), $u^iv^j \uparrow v^ju^i$ contradicting the fact that X is a pcode. ■

The next proposition implies that $p_e, p_c \in F$.

Proposition 8. *Let X be a nonempty subset of $W\backslash\{\varepsilon\}$. If X is a pcode, then X is an p_c-antichain* (*respectively*, *p_e-antichain*).

Proof. Let X be a pcode. The case where X contains only one partial word is trivial. Using Proposition 1 and Lemma 8, it is enough to show the result for p_c. Let $u, v \in X$ be such that $u \neq v$ and up_cv. Then $v \subset ux$, $v \subset xu$ for some $x \in W_0$. If $x = \varepsilon$, then $v \subset u$. This gives $v \uparrow u$ and hence $uv \uparrow vu$. If $x \neq \varepsilon$, then $uv \subset uxu$, $vu \subset uxu$ and so $uv \uparrow vu$. In either case we get a contradiction with the fact that X is a pcode. Hence X is an antichain with respect to Pc. ■

The above proposition does not hold for p_o. For example, $X = \{ab^2, ba, ab, b^2a\}$ is an p_o-antichain but not a pcode since $(ab^2)(ba) = (ab)(b^2a)$.

The next two propositions relate two-element pcodes with the relation $u \bigcup_{p\in F} p$.

Proposition 9. *Let $u,v \in W\backslash\{\varepsilon\}$ be such that $|u| < |v|$. Then $u \bigcup_{p\in F} pv$ if and only if $\{u, v\}$ is not a pcode.*

Proof. The condition is obviously necessary. To see that the condition is sufficient, suppose that $\{u, v\}$ is not a pcode and let $(u, v) \notin u \bigcup_{p\in F} p$. Let $y = \{(u, v)\} \cup u \bigcup_{p\in F} p$. Then $u \bigcup_{p\in F} p \sqsubset y$ and $y \in F$, a contradiction. ■

A subset X of W is called *pairwise noncompatible* if $u \not\uparrow v$ for all distinct $u, v \in X$.

Proposition 10. *Let $X \subset W\backslash\{\varepsilon\}$ be pairwise noncompatible. Then X is an $u \bigcup_{p\in F} p$-antichain if and only if for all $u, v \in X$ such that $u \neq v$, $\{u, v\}$ is a pcode.*

Proof. First, suppose that X is an $u \bigcup_{p\in F} p$-antichain. Let $u, v \in X$ be such that $u \neq v$. Without loss of generality, we can assume that $|u| \leq |v|$. Since X is an $u \bigcup_{p\in F} p$-antichain, we have $(u, v) \notin u \bigcup_{p\in F} p$. If $|u| < |v|$, then $\{u, v\}$ is a pcode by Proposition 9. If $|u| = |v|$, then $u \not\uparrow v$ since X is pairwise noncompatible. Certainly, in this case, $\{u, v\}$ is a pcode.

Conversely, suppose to the contrary that there exist $u, v \in X$ such that $u \neq v$ and $(u , v) \in u \bigcup_{p\in F} p$. The set $\{u, v\}$ is a pcode by our assumption. Since $u \bigcup_{p\in F} p$ is strict, we have $|u| \leq |v|$. If $|u| < |v|$, then $\{u, v\}$ is not a pcode by Proposition 9, a contradiction. If $|u| = |v|$, then $v \sqsubset u$ since $u \bigcup_{p\in F} p$ is strict. So $u \uparrow v$ contradicting the fact that $\{u, v\}$ is a pcode. So X is an $u \bigcup_{p\in F} p$-antichain. ■

4.2. The class G

We now consider the following class of binary relations on W partially ordered by inclusion:

$G = \{p \mid p$ is a strict binary relation on W such that every antichain with respect to p is a pcode$\}$.

The class G was considered in [19] for strict positive binary relations on W_0.

The following proposition gives a closure property for G and immediately implies that G is closed under union.

Proposition 11. *Let y be a strict binary relation on $W$and let $p \in G$. If $(p) \subset (y)$, then $y \in G$.*

Proposition 12. *Let $u \in W_0\backslash\{\varepsilon\}$, $v \in W\backslash\{\varepsilon\}$ be such that $|u| \leq |v|$. If $\{u, v\}$ is an antichain with respect to p_o* (*respectively*, *p_p, p_s, p_f, p_d, p_l*), *then $\{u, v\}$ is a pcode.*

Proof. By Proposition 1 and Lemma 8, it is enough to show the result for p_o. Suppose to the contrary that $\{u, v\}$ is not a pcode. Then there exist an integer $n \geq 1$ and partial words $u_1,...,u_n,v_1,...,vn \in \{u , v\}$ such that

$u_1u_2...u_n \uparrow v_1v_2...v_n$,

and with $|u_1u_2...u_n|$ as small as possible contradicting Proposition 3. We hence have $u_1 \neq v_1$ and $u_n \neq v_n$. If $n = 1$, then $u \uparrow v$. Since u is full, we get $v \subset u$ and so up_ov, which is a contradiction. So we may assume that $n \geq 2$. There are four possibilities: $u_1 = u_n = u$, $v_1 = v_n = v$; $u_1 = v_n = u$, $v_1 = u_n = v$; $u_1 = v_n = v$, $v_1 = u_n = u$; and $u\ 1 = u_n = v$, $v_1 = v_n = u$. In all cases, put $u_2...u_{n-1} = x$ and $v_2...v_{n-1} = y$. These possibilities can be rewritten as

(1) $uxu \uparrow vyv$,
(2) $uxv \uparrow vyu$,
(3) $vxu \uparrow uyv$,
(4) $vxv \uparrow uyu$.

If $|u| = |v|$, for any of possibilities (1)–(4) we have $u \uparrow v$ which leads to a contradiction. If $|u| < |v|$, for any of possibilities (1)–(4) there exist $w, w', z, z' \in W\backslash\{\varepsilon\}$ such that $v = wz = z'w'$, $w \uparrow u$, and $w' \uparrow u$. The latter two relations give $w \subset u$ and $w' \subset u$ since u is full. There exist $z_1, z_2 \in W_0$ such that $z \subset z_1$ and $z' \subset z_2$. We get $v = wz \subset uz \subset uz_1$, $v = z'\,w' \subset z'\,u \subset z_2u$ and so up_ov, which is a contradiction. ■

The converse of the above proposition is not true. For example, the set $X = \{a, aba\}$ is a pcode, but $a\ p_oaba$. The above proposition is not true if u has a hole. The set $\{u, v\}$ where $u = a\Diamond$ and $v = \Diamond a$ is an p_l-antichain, but $\{u, v\}$ is not a pcode. This latter example shows that $p_e, p_c, p_o, p_p, p_s, p_f, p_d$, and p_l are not in G.

5. Prefix and suffix orderings

In this section, we discuss the prefix and the suffix orderings which we denote by $\preceq p$ and $\preceq s$ instead of p_p and p_s.

It is well-known that a subset X of $W_0\backslash\{\varepsilon\}$ is an antichain with respect to $\preceq_p$ if and only if X is a *prefix code*, or if for any $u \in X$, $ux \notin X$ for all $x \in W_0\backslash\{\varepsilon\}$ [18].

We now show that with partial words, the antichains with respect to $\preceq_p$ are the *anti-prefix* sets defined as follows.

Definition 6. Let $X \subset W\backslash\{\varepsilon\}$. Then X is *anti-prefix* if for any $u \in X$, the following conditions hold:

- If $v \sqsubset u$, then $v \notin X$.
- If $v \subset ux$ for some $x \in W0\backslash\{\varepsilon\}$, then $v \notin X$.

It is immediate that a singleton set is anti-prefix and any nonempty subset of an anti-prefix set is anti-prefix. Hence any nonempty intersection of anti-prefix sets is anti-prefix.

Proposition 13. *Let $X \subset W_0\backslash\{\varepsilon\}$. Then X is an antichain with respect to $\preceq_p$ if and only if X is anti-prefix.*

Proof. Assume that X is an antichain with respect to $\preceq_p$. Let $u \in X$, and suppose to the contrary that X is not anti-prefix. So either there exists $v \in X$ with $v \sqsubset u$, or there exist $v \in X$ and $x \in W_0\backslash\{\varepsilon\}$ such that $v \subset ux$. In either case, we have $u, v \in X$, $u \neq v$, and $u \preceq_p v$ contradicting our assumption. On the other hand, if X is anti-prefix, then suppose to the contrary that there exist $u, v \in X$ with $u \neq v$ and $u \preceq_p v$. Then $v \sqsubset u$ or there exists $x \in W_0\backslash\{\varepsilon\}$ such that $v \subset ux$. In either case, $v \notin X$ a contradiction. ■

Corollary 1. *Let $u \in W_0\backslash\{\varepsilon\}$, $v \in W\backslash\{\varepsilon\}$ be such that $|u| \leq |v|$. If $\{u, v\}$ is anti-prefix, then $\{u, v\}$ is a pcode.*

Proof. The result follows from Propositions 12 and 13. ■

A subset X of $W_0\backslash\{\varepsilon\}$ is an antichain with respect to $\preceq_s$ if and only if X is a *suffix code*, or if for any $u \in X$, $xu \notin X$ for all $x \in W_0\backslash\{\varepsilon\}$ [18].

The family of anti-suffix sets coincides with the family of antichains with respect to $\prec_s$.

Definition 7. Let $X \subset W\backslash\{\varepsilon\}$. Then X is *anti-suffix* if for any $u \in X$, the following conditions hold:

- If $v \sqsubset u$, then $v \notin X$.
- If $v \subset xu$ for some $x \in W_0\backslash\{\varepsilon\}$, then $v \notin X$.

Proposition 14. *Let $X \subset W\backslash\{\varepsilon\}$. Then X is an antichain with respect to $\preceq_s$ if and only if X is anti-suffix.*

Proof. The proof is similar to that of Proposition 13. ■

Corollary 2. *Let $u \in W_0\backslash\{\varepsilon\}$, $v \in W\backslash\{\varepsilon\}$ be such that $|u| \leq |v|$. If $\{u, v\}$ is anti-suffix, then $\{u, v\}$ is a pcode.*

Proof. The result follows from Propositions 12 and 14. ■

We end this section by noticing that there exist anti-prefix (or anti-suffix) sets that are not pcodes. For example, the set $\{u, v\}$ where $u = a\Diamond b$ and $v = abbaab$ is both anti-prefix and anti-suffix, but $\{u, v\}$ is not a pcode since $u^2 \uparrow v$.

6. Border ordering

In this section, we discuss the border ordering which we denote by $\preceq_o$ instead of p_o. Let v be a nonempty partial word. By definition, $\varepsilon \prec_o v$ and $v \not\subset \varepsilon$, and let $N(v)$ be the number of partial words u satisfying $u \prec_o v$ and $v \not\subset u$. For any integer $i \geq 0$, define O_i as follows:

$O_0 = \{\varepsilon\}$

and for $i \geq 1$,

$O_i = \{v \mid v \in W\backslash\{\varepsilon\} \text{ and } N(v) = i\}$.

We are particularly interested in the partial words in O_1. A nonempty partial word v is called *unbordered* if $u \preceq_o v$ for some nonempty partial word u implies $v \subset u$. Clearly, v is unbordered if $v \subset ux$ and $v \subset yu$ imply $x = y = \varepsilon$ or $u = \varepsilon$. The fact that v is unbordered means that there exist no nonempty partial words u, x, y satisfying $v \subset ux$ and $v \subset yu$. Note that O_1 is the set of all nonempty unbordered partial words, which is a subset of the primitive partial words [5]. From the point of view of the partial order $\preceq_o$, we call the partial words in O_1 *o-primitive*. It is easy to see that $W = \bigcup_{i\geq 0} O_i$ with $O_i \cap O_j = \emptyset$ if $i \neq j$.

The following extend results of [2].

Proposition 15. *Let $u \in W\backslash\{\varepsilon\}$ be such that $0 \notin H(u)$. If $||A|| \geq 2$, then there exists $v \in W_0$ such that uv is unbordered.*

Proof. Let a be the first letter of u, and let $b \in A\backslash \{a\}$. We claim that the partial word $w = uab^{|u|}$ is unbordered. To see this, suppose there exist nonempty partial words x, y, z satisfying $w \subset xy$, $w \subset zx$. Since $w \subset xy$, the nonempty word x starts with the letter a. Since $w \subset zx$, we have $|x| > |u|$. But then we have $x = x'ab^{|u|}$ for some $x' \in W$, and also $x = u'ab^{|x'|}$ for some $u' \in W$ satisfying $|u'| = |u|$. Thus $|x'| = |u|$, and hence $w \subset x$, a contradiction. ■

Proposition 16. *Let $X \subset W\backslash\{\varepsilon\}$ be a pcode. If $u \in W0$ is an unbordered word such that $u \notin F(C(X^*))$, then the set $Y = X \cup \{u\}$ is apcode.*

Proof. Let $U = W\backslash C(WuW)$. Then by assumption $X^* \subset U$. Let us first observe the following property of the set $V = Uu$: for all $v, v' \in U$, $v'u \uparrow vux$ for some x implies $v \uparrow v'$. To see this, suppose that $v'u \uparrow vux$ for two partial words v and v' in U and some x. If $|vu| > |v'|$, then $vu \uparrow v'y$ with $u = yz$ for some y, z. We deduce that $yz \uparrow z'y$ for some z'. If $z = \varepsilon$, then $vu \uparrow v'u$ and $v \uparrow v'$. If $z \neq \varepsilon$, then since y is full, by conjugacy on partial words [9], there exist words x', y' such that $z' \subset x'y'$, $z \subset y'x'$, and $y \subset (x'y')^n x'$ for some integer $n \geq 0$. But then $u \subset (x'y')^{n+1}x'$, and since u is unbordered, $x' = \varepsilon$. If $n > 0$, u is bordered, and if $n = 0$, we get $y = \varepsilon$ and so $vu \uparrow v'$. This leads to $v' \in C(WuW)$, which is a contradiction. Hence $|vu| \leq |v'|$, and $vuy \uparrow v'$ for some y. But then again v' is in $C(WuW)$, a contradiction.

Now, we show that Y is a pcode. Assume the contrary and consider a relation

$u_1u_2...u_m \uparrow v_1v_2...v_n$

with $u_1,...,u_m,v_1,...,v_n \in Y$, and $u_1 \neq v_1$. The set X being a pcode, one of these partial words must be u. Assume that one of $u_1,...,u_m$ is u, and let i be the smallest index such that $u_i = u$. Since $WuW \cap C(X^*) = \emptyset$, it follows that $Wu_i\, W \cap C(X^*) = \emptyset$. Consequently one of $v_1,...,v_n$ is u. Let j be the smallest index such that $v_j = u$. Then $u_1...u_{i-1}u$, $v_1...v_{j-1}u \in V$ whence $u_1...u_{i-1} \uparrow v_1...v_{j-1}$ by the abovementioned property of V. The set X is a pcode, thus from $u_1 \neq v_1$ it follows that $i = j = 1$ leading to a contradiction. ■

A pcode X is called *maximal* over A if it is not a proper subset of any other pcode over A. It is called *complete* if $F(C(X^*)) = W$.

Theorem 2. *Let $X \subset W\backslash\{\varepsilon\}$. If X is a maximal pcode, then X is complete.*

Proof. Let $X \subset W\backslash\{\varepsilon\}$ be a maximal pcode that is not complete. If $||A|| = 1$, then $X = \emptyset$ and X is not maximal. If $||A|| \geq 2$, consider a word $u \in W\backslash\{\varepsilon\}$ such that $u \notin F(C(X^*))$. We may choose u in W_0. According to Proposition 15, there exists a word $v \in W_0$ such that uv is unbordered. We have $uv \notin F(C(X^*))$, and it then follows from Proposition 16 that $X \cup \{uv\}$ is a pcode. Thus X is not maximal, a contradiction. ■

7. Commutative ordering

In this section, we discuss the commutative ordering that we denote by $\preceq_c$ instead of p_c.

Lemma 9. *Let $u, v \in W\backslash\{\varepsilon\}$ be such that v is non $\{|u|, |v| - |u|\}$-special. Then $u \preceq_c v$ if and only if there exists a primitive word z and integers m, n such that $u \subset z^m$ and $v \subset uz^n \subset z^{m+n}$, $v \subset z^n u \subset z^{m+n}$.*

Proof. Let $u, v \in W\backslash\{\varepsilon\}$ be such that v is non $\{|u|, |v| - |u|\}$-special. If $u \preceq_c v$, then for some word $x \in W_0$, we have $v \subset xu$, $v \subset ux$. Let $u' \in W_0$ be such that $u \subset u'$. If $x = \varepsilon$, then $v \subset u \subset u'$ and there exists a primitive word z and a positive integer m such that $u' = z^m$. Hence $u \subset z^m$, $v \subset uz^0 \subset z^{m+0}$, $v \subset z^0u \subset z^{m+0}$ and the result follows. So we may assume that x is nonempty. We get $v \subset xu'$, $v \subset u'x$ and thus by Lemma 3, $xu' = u'x$. By Lemma 2, there exists a primitive word z and positive integers m, n such that $u' = z^m$ and $x = z^n$ ($z = \sqrt{x}$). This in turn implies that $u \subset u' \subset z^m$ and $v \subset ux = uz^n \subset z^{m+n}$, $v \subset xu = z^nu \subset z^{m+n}$. ■

It is known that a subset X of $W_0\backslash\{\varepsilon\}$ is an antichain with respect to $\preceq_c$ if and only if X is *anti-commutative*, or if for all $u, v \in X$ satisfying $u \neq v$, we have $uv \neq vu$ [15]. Now, we call a subset X of $W\backslash\{\varepsilon\}$ *anti-commutative* if for all $u, v \in X$ satisfying $u \neq v$, we have $uv \not\uparrow vu$. Certainly, every pcode is anti-commutative.

Proposition 17. *Let $X \subset W\backslash\{\varepsilon\}$ be pairwise nonspecial. If X is anti-commutative, then X is an antichain with respect to $\preceq_c$.*

Proof. If X is anti-commutative, then let us show that X is an antichain with respect to $\preceq_c$. Suppose to the contrary that there exist $u, v \in X$ with $u \neq v$ and $u \preceq_c v$. The latter implies that $|u| \leq |v|$. By assumption, v is non $\{|u|, |v| - |u|\}$-special, and by Lemma 9, there exists a primitive word z and integers m, n such that $u \subset z^m$ and $v \subset z^{m+n}$. But then $uv \uparrow vu$ contradicting the fact that X is anti-commutative. ■

Proposition 18. *Let $X \subset W \backslash \{\varepsilon\}$. Let $u, v \in X$ be such that u is full, $u \neq v$, and uv is non $\{|u|, |v|\}$-special. If X is an antichain with respect to $\preceq_c$, then $uv \not\uparrow vu$.*

Proof. Suppose to the contrary that $uv \uparrow vu$. There exists a word z such that $uv \subset z$ and $vu \subset z$. Put $z = xy$ where $u \subset x$ and $v \subset y$. We have $uv \subset xy$, and by Lemma 4 we also have $uv \subset yx$. Lemma 3 implies $xy = yx$, and so x, y are powers of a common word. Say $x = w^m$ and $y = w^n$ for some word w and integers m, n. Since u is full, we have $u = w^m$. If $m = n$, then $v \subset y = w^n = w^m = u$, and so $u \preceq_c v$. For the case $m < n$, we have $v \subset y = w^n = uw^{n-m} = w^{n-m}u$ and thus $u \preceq_c v$. Similarly, we can show that if $m > n$, then $v \preceq_c u$. In all cases, we obtain a contradiction.

In Proposition 18, both the assumptions that u is full and uv is non $\{|u|, |v|\}$-special are needed. Indeed, if we put $X = \{u, v\}$ where $u = a\Diamond b$ and $v = aab\Diamond ab$, we get that X is an antichain with respect to $\preceq_c$ and that $uv \uparrow vu$. This example is such that u is nonfull and uv is non $\{|u|, |v|\}$-special. Now, if we put $X = \{u, v\}$ where $u = abbaab$ and $v = \Diamond\Diamond\Diamond\Diamond$, we get that X is an antichain with respect to $\preceq_c$ and that $uv \uparrow vu$. This example is such that u is full and $u\,v$ is $\{|u|, |v|\}$-special.

Let $u', x, y \in W_0 \backslash \{\varepsilon\}$, $v' \in W \backslash \{\varepsilon\}$ be such that $|x| = |y|$ and $|u'x| = |v'|$. Then the set $\{u, v\}$ where $u = u'x$ and $v = v'y$ (respectively, $u = xu'$ and $v = yv'$) is said to be of *type* 1 (respectively, *type 2*) if v is not $\{|u|, |x|\}$-special.

Proposition 19. *Let $u, v \in W \backslash \{\varepsilon\}$ be such that $\{u, v\}$ is of type* 1 *or type* 2. *Then $uv \not\uparrow vu$ if and only if $\{u, v\}$ is a pcode.*

Proof. We prove the result for type 1 (type 2 is similar). If $\{u, v\}$ is a pcode, then clearly $uv \not\uparrow vu$. Conversely, assume that $\{u, v\}$ is not a pcode and $uv \not\uparrow vu$. Then there exist an integer $n \geq 1$ and partial words $u_1, \ldots, u_n$, $v_1, \ldots, v_n \in \{u, v\}$ such that

$$u_1 u_2 \ldots u_n \uparrow v_1 v_2 \ldots v_n$$

and with $|u_1 u_2 \ldots u_n|$ as small as possible contradicting Proposition 3. We hence have $u_1 \neq v_1$ and $u_n \neq v_n$, and we may assume that $n > 2$. There are the four possibilities (1)–(4) as in Proposition 12. Since $\{u, v\}$ is of type 1, there exist $u', z_1, z_2 \in W_0 \backslash \{\varepsilon\}$, $v' \in W \backslash \{\varepsilon\}$ such that $|z_1| = |z_2|$, $|u'z_1| = |v'|$, $u = u'z_1$, $v = v'z_2$, and v is not $\{|u|, |z_1|\}$-special. Any possibility gives $v' \uparrow u$. Substituting u by $u'z_1$ and v by $v'z_2$ in (1)–(4) we get

(5) $u'z1xu'z1 \uparrow \uparrow v'z2yv'z2$,
(6) $u'z1xv'z2 \uparrow \uparrow v'z2yu'z1$,
(7) $v'z2xu'z1 \uparrow \uparrow u'z1yv'z2$,
(8) $v'z2xv'z2 \uparrow \uparrow u'z1yu'z1$.

Any possibility implies $z_1 \uparrow z_2$, and hence $z_1 = z_2$ since both z_1 and z_2 are full. So $v = v'z_1$, and hence both u and v end with z_1, and the same is true for both x and y. We deduce that $v \uparrow z_1 u$, and so $v'z_1 \uparrow z_1 u$ and hence $v'z_1 \subset z_1 u$. The fact that $v' \uparrow u$ implies $v'z_1 \subset uz_1$. By Lemma 3, we get $uz_1 = z_1 u$ since v is not $\{|u|, |z_1|\}$-special, and by Lemma 2, u and z_1 are powers of a common word. So $v = v'z_1 \subset uz_1$ is contained in a power of that same common word. But then $uv \uparrow vu$, a contradiction. ■

Note that the above proposition is not true in general. The set $\{u, v\}$ where $u = a\Diamond b$ and $v = abbaab$ satisfies $uv \not\uparrow vu$, but $\{u, v\}$ is not a pcode since $u^2 \uparrow v$.

8. Deciding the pcode property

Here, we give (in Section 8. 1) a brief overview of Head and Weber's domino technique on words [13], and we give (in Section 8.2) our extension of this technique to partial words. As an application, the pcode property turns out to be decidable.

8. 1. Domino technique on words

Let X be a nonempty finite subset of A^+. For $\alpha, \beta \in X^*$ satisfying $\alpha = \beta$ put $\alpha = \alpha_1\alpha_2...\alpha_m$ and $\beta = \beta_1 \beta_2...\beta_n$ for some $\alpha_1,...,a_m, \beta_1,...,\beta_n \in X$. We say that the relation $\alpha = \beta$ is *trivial* if $m = n$ and $\alpha_1 = \beta_1,..., \alpha_m = \beta_m$. We say that the relation $\alpha = \beta$ is *factorizable* if there exist $\alpha', \alpha'', \beta', \beta'' \in X^+$ such that $\alpha = \alpha'\alpha''$, $\beta = \beta'\beta''$ $\alpha' = \beta'$, and $\alpha'' = \beta''$.

In order to study the relations satisfied by X, Guzmán suggested to look at the *simplified domino graph* and the *domino function* of X [11] (this approach was further considered in [3] for instance). The simplified domino graph of X is a subgraph of the Head and Weber's domino graph of X defined in [13].

Let *Prefix*(X) be the set of all prefixes of words in X, and let $G = (V, E)$ be the directed graph with vertex set

$V = \{open, close, \binom{u}{\varepsilon}, \binom{\varepsilon}{u} \mid u \in Prefix\ (X)\backslash\{\varepsilon\}\}$

and with edge set $E = E1 \cup E2 \cup E3 \cup E4$ where

$E_1 = \{(open, \binom{\varepsilon}{u}) \mid u \in X\}$,
$E_2 = \{(\binom{u}{\varepsilon}, close) \mid u \in X\}$,
$E_3 = \{(\binom{u}{\varepsilon}, \binom{uv}{\varepsilon}), (\binom{\varepsilon}{u}, \binom{\varepsilon}{uv}) \mid v \in X\}$,
$E_4 =\{(\binom{u}{\varepsilon}, \binom{\varepsilon}{v}), (\binom{\varepsilon}{u}, \binom{v}{\varepsilon}) \mid uv \in X\}$.

The simplified domino graph associated with X is the directed graph $G' = (V', E')$ where V' consists of *open*, *close* and those vertices v in V such that there exists a path from *open* to *close* that goes through v, and E' consists of those edges e in E such that there exists a path from *open* to *close* going through e. The simplified domino graph of X will be denoted by $G\ (X)$. The domino function associated with X is the mapping d from E to $\{\binom{u}{\varepsilon}, \binom{\varepsilon}{u} \mid u \in X\}$ defined on

E_1 by (open,$(\binom{\varepsilon}{u}) \mapsto \binom{u}{\varepsilon}$,
E_2 by $(\binom{u}{\varepsilon}$, close) $\mapsto \binom{u}{\varepsilon}$,
E_3 by $(\binom{u}{\varepsilon}, \binom{uv}{\varepsilon}) \mapsto \binom{\varepsilon}{v}$ and $(\binom{\varepsilon}{u}, \binom{\varepsilon}{uv}) \mapsto \binom{u}{\varepsilon}$,
E_4 by $(\binom{u}{\varepsilon}, \binom{\varepsilon}{v}) \mapsto \binom{uv}{\varepsilon}$ and $(\binom{\varepsilon}{u}, \binom{v}{\varepsilon}) \mapsto \binom{\varepsilon}{uv}$.

The *domino* associated with an edge e of E is the domino $d(e) = \binom{d_1(e)}{d_2(e)}$ The function d induces mappings d_1 and d_2 from E to $X \cup \{\varepsilon\}$ also called domino functions. If $p = e_1e_2...e_i$ is a path in G, then $d(e_1)d(e_2)...d(e_i)$ (respectively, $d_1(e_1)d_1(e_2)...d_1(e_i)$, $d_2(e_1)d_2(e_2)...d_2(e_i)$) is denoted by $d\ (p)$ (respectively, $d_1(p)$, $d_2(p)$).

A path p in $G(X)$ from *open* to some vertex $\binom{u}{\varepsilon}$ (respectively, $\binom{\varepsilon}{u}$) is trying to find two decodings of the same message over X into codewords beginning with distinct codewords. The decodings obtained so far are $d_1(p)$ and $d_2(p)$. The word u in A^* denotes the back-log of the first (respectively, second) decoding as against the second (respectively, first) one.

The next proposition illustrates how the paths from open to close in $G\ (X)$ correspond to nontrivial nonfactorizable relations satisfied by X.

Proposition 20 (*Guzmán [11]*). *Let X be a nonempty finite subset of* A^+. *For* $\alpha, \beta \in X^*$, $\alpha = \beta$ *is a nontrivial nonfactorizable relation if and only if there exists a path p in G (X) from open to close such that* $d(p) = \binom{\alpha}{\beta}$ *or* $d(p) = \binom{\beta}{\alpha}$.

If $G(X)$ is treated as an automaton with initial state *open* and final state *close*, the set accepted by $G(X)$ consists of dominoes $\binom{\alpha}{\beta}$ such that $\alpha, \beta \in X^*$ and $\alpha = \beta$.

The code property of X can be characterized in terms of its simplified domino graph $G(X)$ as follows.

Theorem 3 (*Head and Weber [13]*). *Let X be a nonempty finite subset of* A^+. *Then X is a code if and only if there is no path in G (X) from open to close.*

8.2. Domino technique on partial words

In this section, we show that it is decidable whether or not a nonempty finite subset of $W \backslash \{\varepsilon\}$ is a pcode. Our approach is based on an adaptation of the domino technique of the previous section.

Let X be a nonempty finite subset of $W \backslash \{\varepsilon\}$. For $\alpha, \beta \in X^*$ satisfying $\alpha \uparrow \beta$, put $\alpha = \alpha_1 \alpha_2 ... \alpha_m$ and $\beta = \beta_1\beta_2...\beta_n$ for some $\alpha_1,...,\alpha_m, \beta_1,...,\beta_n \in X$. We say that the relation $\alpha \uparrow \beta$ is *trivial* if $m = n$ and $\alpha_1 = \beta_1,...,\alpha_m = \beta_m$. We say that the relation $\alpha \uparrow \beta$ is *factorizable* if there exist $\alpha', \alpha'', \beta', \beta'' \in X^+$ such that $\alpha = \alpha'\alpha''$, $\beta = \beta'\beta''$, $\alpha' \uparrow \beta'$, and $\alpha'' \uparrow \beta''$.

In order to study the compatibility relations

$\alpha_1\alpha_2 ... \alpha_m \uparrow \beta_1\beta_2 ... \beta_n$

where $\alpha 1 ,... , \alpha_m, \beta_1,...,\beta_n \in X$, we extend the technique of Section 8.1. Let $Prefix(X)$ be the set of all prefixes of partial words in X, and let $G = (V, E)$ be the directed graph with vertex set

$V = \{open, close, \binom{u}{\varepsilon}, \binom{\varepsilon}{u} \mid u \in C(Prefix(X)) \backslash \{\varepsilon\}\}$

and with edge set $E = E_1 \cup E_2 \cup E_3 \cup E_4$ where

$E_1 = \{(open, \binom{\varepsilon}{u}) \mid u \in X\}$,
$E_2 = \{(\binom{u}{\varepsilon}, close), (\binom{\varepsilon}{u}, close) \mid u \in C(X)\}$,
$E_3 = \{(\binom{u}{\varepsilon}, \binom{uv}{\varepsilon}), (\binom{\varepsilon}{u}, \binom{\varepsilon}{uv}) \mid v \in X\}$,
$E_4 = \{\binom{u}{\varepsilon}, \binom{\varepsilon}{v}), (\binom{\varepsilon}{u}, \binom{v}{\varepsilon}) \mid w = u'v, u \uparrow u', w \in X\}$.

The *simplified domino graph* associated with X is the directed graph $G' = (V', E')$ where V' consists of *open*, *close* and those vertices v in V such that there exists a path from *open* to *close* that goes through v, and E' consists of those edges e in E such that there exists a path from *open* to *close* going through e. The simplified domino graph of X will be denoted by $G(X)$. The *domino function* associated with X is the mapping d from E to the set of nonempty subsets of $\{(\binom{u}{\varepsilon}, \binom{\varepsilon}{v}) \mid u \in X\}$ defined on

E_1 by $(open, \binom{\varepsilon}{u}) \mapsto \{\binom{u}{\varepsilon}\}$,
E_2 by $(\binom{u}{\varepsilon}, close) \mapsto \{\binom{v}{\varepsilon} \mid u \uparrow v$ and $v \in X\}$ and $(\binom{\varepsilon}{u}, \text{close}) \mapsto \{\binom{\varepsilon}{v} \mid u \uparrow v$ and $v \in X\}$
E_3 by $(\binom{u}{\varepsilon}, \binom{uv}{\varepsilon}) \mapsto \binom{v}{\varepsilon}$ and $(\binom{\varepsilon}{u}, \binom{\varepsilon}{uv}) \mapsto \{\binom{v}{\varepsilon}\}$,
E_4 by $(\binom{u}{\varepsilon}, \binom{\varepsilon}{v}) \mapsto \{\binom{w}{\varepsilon} \mid w = u'u\ u \uparrow u'$ and $w \in X\}$ and $(\binom{\varepsilon}{u}, \binom{v}{\varepsilon}) \mapsto \{\binom{\varepsilon}{w} \mid w = u'v, u \uparrow u'$ and $w \in X\}$

The *domino set* associated with an edge e of E is the set $d(e)$. If $p = e_1e_2...e_i$ is a path in G, the set

$d(e_1)d(e_2) \ldots d(e_i) = \{x_1x_2 \ldots x_i \mid x_1 \in d(e_1), x_2 \in d(e_2), \ldots, x_i \in d(e_i)\}$

is denoted by $d(p)$. For $x = \binom{y_1}{z_1}\binom{y_2}{z_2}\ldots\binom{y_i}{z_i}$ in $d(p)$, we abbreviate $y_1y_2\ldots y_i$ by *above*(x) and $z_1z_2\ldots z_i$ by *below*(x). We will also write $x = \binom{above(x)}{below(x)}$ Note that *above*(x), (*below*(x) are in X^*.

A path p in $G(X)$ from *open* to some vertex $\binom{u}{\varepsilon}$ is trying to find a nontrivial compatibility relation over X. The factorizations obtained so far for a particular $x \in d(p)$ are *above*(x) and *below*(x). More precisely, if *above*(x) = $\alpha_1\ \alpha_2 \ldots \alpha_m$ and *below*(x) = $\beta_1\beta_2 \ldots \beta_n$, then $\alpha_1 \neq \beta_1$ and $\alpha_1\ \alpha_2 \ldots \alpha_m u \uparrow \beta_1\beta_2\ldots\beta_n$ and u is a suffix of $\beta_1\beta_2\ldots\beta_n$. The partial word u denotes the backlog of the first factorization as against the second one. Similarly, if p is from *open* to some vertex $\binom{\varepsilon}{u}$, then $\alpha_1 \neq \beta_1$ and $\alpha_1\ \alpha_2 \ldots \alpha_m \uparrow \beta_1\beta_2 \ldots \beta_n u$ and u is a suffix of $\alpha_1\alpha_2\ldots\alpha_m$. In this case, u denotes the backlog of the second factorization as against the first one.

In the sequel, in order to simplify the notation, we identify both *open* and *close* with $\binom{\varepsilon}{\varepsilon}$.

Lemma 10. 1. *If $u \in C$ (Prefix(X)) and there exists a path p in G (X) from open to* $\binom{u}{\varepsilon}$ (*respectively,* $\binom{\varepsilon}{u}$), *then* $d(p)$ *consists of elements of the form* $\binom{\alpha}{\beta u}$ (*respectively,* $\binom{\alpha u}{\beta}$) *for some $\alpha, \beta \in W$ satisfying $\alpha \uparrow \beta$.*
2. *If there exists a path p in G (X) from open to close such that* $\binom{\alpha}{\beta} \in d(p)$, *then $\alpha \uparrow \beta$ is β a nonfactorizable compatibility relation satisfied by X. Moreover, if p is of length at least* 3, *then $\alpha \uparrow \beta$ is nontrivial.*

Proof. First, Statement 1 follows by induction. The only path of length 1 from *open* is an E_1-edge of the form (*open*, $\binom{\varepsilon}{u}$) for some $u \in X$. Here, $d(p) = \{\binom{u}{\varepsilon}\}$ and the result follows with $\alpha = \beta = \varepsilon$. Now, consider the path $q = pe$ where p is a path from *open* to $\binom{u}{\varepsilon}$ and e is an edge from $\binom{u}{\varepsilon}$. By the inductive hypothesis, $d(p)$ consists of elements of the form $\binom{\alpha}{\beta u}$ for some $\alpha, \beta \in W$ satisfying $\alpha \uparrow \beta$. For $e = (\binom{u}{\varepsilon}, close) \in E_2$, $d(pe) = d(p)d(e)\ \beta u$ consists of elements of the form $\binom{\alpha}{\beta u}\binom{\varepsilon}{v} = \binom{\alpha v}{\beta u} = \binom{\alpha'}{\beta'}$ where $u \uparrow v$ and $v \in X$. For $e = (\binom{u}{\varepsilon}, \binom{uv}{\varepsilon}) \in E_3$, $d(pe)$ consists of elements of the form $\binom{\alpha}{\beta u}\binom{\varepsilon}{v} = \binom{\alpha}{\beta uv} = \binom{\alpha'}{\beta'\mathrm{uv}}$ where $v \in X$. Finally, for $e = (\binom{u}{\varepsilon}, \binom{\varepsilon}{v}) \in E_4$, $d(pe)$ consists of elements of the form $\binom{\alpha}{\beta u}\binom{w}{\varepsilon} = \binom{\alpha w}{\beta u} = \binom{\alpha u' v}{\beta \mathrm{u}} = \binom{\alpha' v}{\beta'}$ where $w = u'v\ u \uparrow u'$ and $w \in X$. In any case, the result follows with some $\alpha', \beta' \in W$ satisfying $\alpha' \uparrow \beta'$. The result follows similarly when p is a path from *open* to $\binom{\varepsilon}{u}$ and e is an edge from $\binom{\varepsilon}{u}$

Second, let us show that Statement 2 holds. If there exists a path p from *open* to *close* such that $\binom{\alpha}{\beta} \in d(p)$, then by Statement 1, $\alpha \uparrow \beta$ since *close* = $\binom{\varepsilon}{\varepsilon}$. But by the definition of $d(p)$, we have $\alpha, \beta \in X^*$ and thus $\alpha \uparrow \beta$ is a compatibility relation satisfied by X. ■

The next lemma shows how to obtain the path corresponding to a given nontrivial non-factorizable compatibility relation. First, we need some definitions.

For two partial words $\alpha, \beta \in W$, we write $\alpha \leq_\uparrow \beta$ if $\alpha \in C(Prefix(\beta))$ where *Prefix*(β) is the set of all prefixes of β, and $\alpha < \beta$ if $\alpha \leq \beta$ and $\alpha \not\uparrow \beta$.

Let $\alpha, \beta \in X^*$, and put $\alpha = \alpha_1\alpha_2\ldots\alpha_m$ and $\beta = \beta_1\beta_2\ldots\beta_n$. We say that $\binom{\alpha}{\beta}$ has a *proper prefix compatibility relation* if there exist $\alpha', \beta' \in X^+$ such that α' is a prefix of α, β' is a prefix of β, $\binom{\alpha}{\beta} \neq \binom{\alpha'}{\beta'}$ and $\alpha' \uparrow \beta'$ is a compatibility relation. Note that a nonfactorizable compatibility relation $\alpha \uparrow \beta$ is such that $\binom{\alpha}{\beta}$ has no proper prefix compatibility relation. We say that $\binom{\alpha}{\beta}$ has the *nppcr* property if the following three conditions hold:

(i) $\alpha \le \beta$ and the suffix y of β satisfying $\beta \uparrow \alpha y$ belongs to $C(Prefix(X))$, or $\beta \le \alpha$ and the suffix y of α satisfying $\alpha \uparrow \beta y$ belongs to $C(Prefix(X))$.
(ii) $\binom{\alpha}{\beta}$ has no proper prefix compatibility relation.
(iii) If $n > 0$, then $m > 0$ and $|\alpha_1| < |\beta_1|$.

Lemma 11. 1. *Let $\alpha, \beta \in X^*$ be such that there exists a path p in $G(X)$ from open to $v_1 \in V$ with $\binom{\alpha}{\beta} \in d(p)$.*

(a) If $v_1 = \binom{u}{\varepsilon}$ and $v \in X$ is such that $uv \in C(Prefix(X))$, then there exist $v_2 \in V$ and a path q from open to v_2 such that $\binom{\alpha}{\beta v} \in d(q)$.

(b) If $v_1 = \binom{u}{\varepsilon}$ and $w = u'v \in X$ is such that $u \uparrow u'$ and $v \in C(Prefix(X))$, then there exist $v_2 \in V$ and a path q from open to v_2 such that $\binom{\alpha w}{\beta} \in d(q)$.

(c) If $v_1 = \binom{\varepsilon}{u}$ and $v \in X$ is such that $uv \in C(Prefix(X))$, then there exist $v_2 \in V$ and a path q from open to v_2 such that $\binom{\alpha v}{\beta} \in d(q)$.

(d) If $v_1 = \binom{\varepsilon}{u}$ and $w = u'v \in X$ is such that $u \uparrow u'$ and $v \in C(Prefix(X))$, then there u exist $v_2 \in V$ and a path q from open to v_2 such that $\binom{\alpha}{\beta w} \in d(q)$.

2. *Let $\alpha, \beta \in X^*$ be such that $\binom{\alpha}{\beta}$ has the nppcr property. Then there exist $v \in V$ and a path|p in $G(X)$ from open to v such that $\binom{\alpha}{\beta} \in d(p)$.*
3. *Let $\alpha, \beta \in X^*$ be such that $\alpha \uparrow \beta$ is a nontrivial nonfactorizable compatibility relation. Then there exists a path p in $G(X)$ from open to close such that $\binom{\alpha}{\beta} \in d(p)$ or $\binom{\beta}{\alpha} \in d(p)$.*

Proof. Cases (a) and (c) of Statement 1 lead to edges in E_3, and Cases (b) and (d) lead to edges in E_2 or E_4 depending on whether $v = \varepsilon$ or $v \neq \varepsilon$. Let us consider Case (b) (the other cases are similar). If $v \neq \varepsilon$, then put $v_2 = \binom{\varepsilon}{v}$ and $e = (\binom{u}{\varepsilon}, \binom{\varepsilon}{v}) \in E_4$. Here, $\binom{\alpha}{\beta}\binom{w}{\varepsilon} = \binom{\alpha w}{\beta} \in d(p)d(e) = d(q)$. On the other hand, if $v = \varepsilon$, then $u' = w$ and take $v_2 = close$ and $e = (\binom{u}{\varepsilon}, close) \in E_2$. Here, $\binom{\alpha}{\beta}\binom{w}{\varepsilon} = \binom{\alpha w}{\beta} \in d(p)d(e) = d(q)$.

For Statement 2, the proof is by induction on $m + n$ where $\alpha = \alpha_1\alpha_2...\alpha_m$ and $\beta = \beta_1\beta_2...\beta_n$. If $m + n = 1$, then by the *nppcr* property, we must have $m = 1$ and $n = 0$. Thus, $\alpha = \alpha_1$ and $\beta = \varepsilon$. Let $v = \binom{\varepsilon}{\alpha_1}$ and p be the path consisting of the edge $e = (open, v) \in E_1$. Then $\binom{\alpha}{\beta} = \binom{\alpha_1}{\varepsilon} \in d(e) = d(p)$.

If $m + n > 1$, then $m > 0$ by the *nppcr* property. So let $\alpha' = \alpha_1\alpha_2...\alpha_{m-1}$, and whenever $n > 0$, let $\beta' = \beta_1\beta_2...\beta_{n-1}$. Note that when $\alpha < \beta$, we have $n > 0$ and β' is defined. Moreover, by the *nppcr* property, we have $\alpha \not\uparrow \beta'$ and $\alpha' \not\uparrow \beta$. So we consider the following cases:

- If $a < \beta$ and $a < \beta'$, then use the inductive hypothesis on $\binom{\alpha}{\beta'}$ and Statement 1(a).
- If $a < \beta$ and $\beta' < a$, then use the inductive hypothesis on $\binom{\alpha}{\beta'}$ and Statement 1(d).
- If $\beta \le a$ and $a' < \beta$, then use the inductive hypothesis on $\binom{\alpha'}{\beta}$ and Statement 1(b).
- If $\beta \le a$ and $\beta < a'$, then use the inductive hypothesis on $\binom{\alpha'}{\beta}$ and Statement 1(c).

Let us consider the third case (the other cases are similar). If $\beta \leq \alpha$ and $\alpha' < \beta$, then put $w = \alpha_m$. Since $\alpha' < \beta$, let u be the suffix of β such that $\beta \uparrow \alpha'u$. The latter and the fact that $\beta \leq \alpha$ imply that $w = u'v \in X$ with $u \uparrow u'$. Since $\beta \leq \alpha$, the suffix v of α satisfying $\alpha \uparrow \beta v$ belongs to $C(Prefix(X))$. We have $u \in C(Prefix(X))$, and so $\binom{\alpha'}{\beta}$ has the *nppcr* property. By the inductive hypothesis, there exist $v_1 \in V$ and a path q from *open* to v_1 such that $\binom{\alpha'}{\beta} \in d(q)$. By Lemma 10(1), $v_1 = \binom{u}{\varepsilon}$. So by Statement (1)(b), there exist $v_2 \in V$ and a path p from *open* to v_2 such that $\binom{\alpha}{\beta} = \binom{\alpha' w}{\beta} \in d(p)$.

For Statement 3, we first note that if α, β are distinct compatible elements of X, then the path $p = e_1e_2$ in $G(X)$ where $e_1 = (open, \binom{\varepsilon}{\alpha})$ and $e_2 = (\binom{\varepsilon}{\alpha}, close)$ is such that $\binom{\alpha}{\beta} \in d(p)$. Otherwise, since $\alpha \uparrow \beta$ is a compatibility relation satisfied by X, Condition (i) of *nppcr* is satisfied. Since it is nonfactorizable, Condition (ii) is satisfied. Finally, since it is nontrivial and nonfactorizable, one of $\binom{\alpha}{\beta}$ and $\binom{\beta}{\alpha}$, say the first, satisfies Condition (iii). Hence $\binom{\alpha}{\beta}$ has the *nppcr* property. By Statement 2, there exist $v \in V$ and a path p from *open* to v such that $\binom{\alpha}{\beta} \in d(p)$. By Lemma 10(1), we must have $v = close$. ■

If $G(X)$ is treated as an automaton with initial state *open* and final state *close*, by Lemma 10(3), the set accepted by $G(X)$ consists of dominoes $\binom{\alpha}{\beta}$ such that $a, \beta \in X^*$ and $a \uparrow \beta$.

A subset X of W containing two distinct compatible partial words is obviously not a pcode. We call X *pairwise noncompatible* if no distinct partial words $u, v \in X$ satisfy $u \uparrow v$. The pcode property of such a set X can be characterized in terms of its simplified domino graph $G(X)$ as follows.

Theorem 4. *Let X be a nonempty finite subset of $W \backslash \{\varepsilon\}$ that is pairwise noncompatible. Then X is a pcode if and only if there is no path of length at least 3 in $G(X)$ from open to close.*

Proof. The above two lemmas illustrate how the paths of length at least 3 from *open* to *close* in $G(X)$ correspond to nontrivial nonfactorizable compatibility relations satisfied by X. Indeed, for $\alpha, \beta \in X^*$, $\alpha \uparrow \beta$ is a nontrivial nonfactorizable compatibility relation if and only if there exists a path p of length at least 3 in $G(X)$ from *open* to *close* such that $\binom{\alpha}{\beta} \in d(p)$ or $\binom{\beta}{\alpha} \in d(p)$. ■

As an example, let us consider the set $X = \{u_1, u_2, u_3, u_4\}$ over the binary alphabet $\{a, b\}$ ($u_1 = a\Diamond b$, $u_2 = aab\Diamond bb$, $u_3 = \Diamond b$, and $u_4 = ba$). The simplified domino graph and function associated with this set are displayed in Fig. 1. The domino set $d(e)$ associated with an edge e of the graph is represented as the label of this edge. Since the domino sets in this example are all singletons, the domino set $\{\binom{u_1}{\varepsilon}\}$ say has been abbreviated by $\binom{u_1}{\varepsilon}$. The reader is invited to take any path of length at least 3 in the simplified domino graph starting at *open* and ending at *close* and to see how a domino sequence x associated with its edges leads to a nontrivial nonfactorizable compatibility relation of the form $above(x) \uparrow below(x)$. The path

$$p = open, \binom{\varepsilon}{a \Diamond b}, \binom{\varepsilon}{a \Diamond b \Diamond b}, \binom{b}{\varepsilon}, \binom{\varepsilon}{b}, \binom{b}{\varepsilon}, \binom{\varepsilon}{b}, \binom{\varepsilon}{a \Diamond b}, close$$

of length at least 3 is from *open* to *close* showing that X is not a pcode. The sequence of labels

$$\binom{u_1}{\varepsilon}\binom{u_3}{\varepsilon}\binom{\varepsilon}{u_2}\binom{u_3}{\varepsilon}\binom{\varepsilon}{u_3}\binom{u_4}{\varepsilon}\binom{u_3}{\varepsilon}\binom{\varepsilon}{u_1}$$

is in $d(p)$ showing that $u_1u_3u_3u_4u_3 \uparrow u_2u_3u_1$ is a nontrivial nonfactorizable compatibility relation over X.

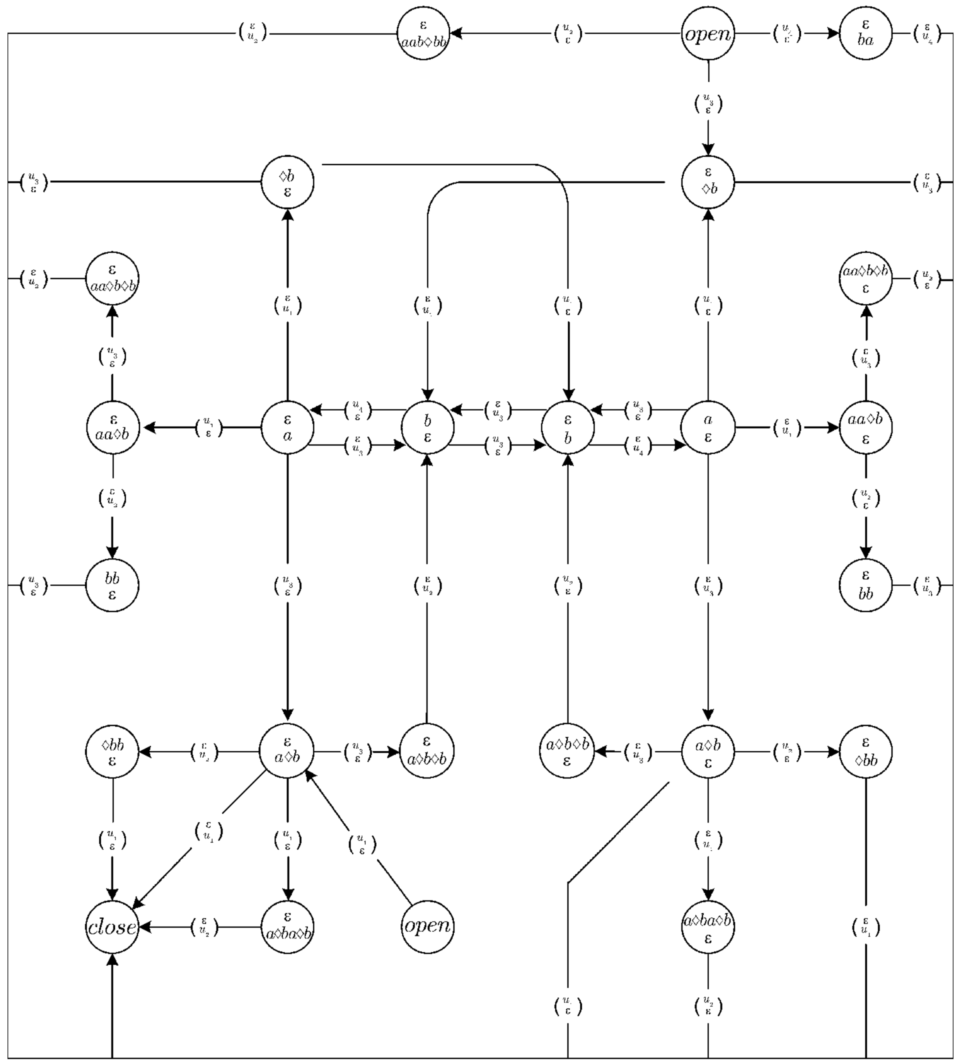

Fig. 1. Simplified domino graph and function (example).

9. Conclusion

In this paper, we have introduced pcodes and have discussed their relation to some partial orderings and to the set of primitive partial words. We have shown the decidability of the pcode property. Open problems abound.

We end this paper with an extension of the class of codes called the *circular codes* to the class of pcodes called the *circular pcodes*.

Let X be a nonempty subset of $W_0 \backslash \{\varepsilon\}$. Then X is called a *circular code* if for all integers $m \geq 1$, $n \geq 1$, words $u_1,\ldots,u_m$, $v_1,\ldots,v_n \in X$, and $r \in W_0$ and $s \in W_0 \backslash \{\varepsilon\}$, the conditions

$$su_2 \ldots u_m r = v_1 v_2 \ldots v_n,$$
$$u_1 = rs$$

imply $m = n$, $r = \varepsilon$, and $u_i = v_i$ for $i = 1,\ldots, m$ [2].

In the case of partial words, we define a *circular pcode* as follows.

Definition 8. Let X be a nonempty subset of $W \backslash \{\varepsilon\}$. Then X is called a *circular pcode* if for all integers $m \geq 1$, $n \geq 1$, partial words $u_1,\ldots,u_m$, $v_1,\ldots,v_n \in X$, and $r \in W$ and $s \in W \backslash \{\varepsilon\}$, the conditions

$$su_2...u_mr \uparrow v_1v_2...v_n,$$
$$u_1 \uparrow rs$$

imply $m = n$, $r = \varepsilon$, and $u_i = v_i$ for $i = 1,...,m$.

It is clear from the definition that a subset X of $W_0\backslash\{\varepsilon\}$ is a circular code if and only if it is a circular pcode. A circular pcode is a pcode, and any subset of a circular pcode is also a circular pcode.

Two partial words u and v are called *conjugate* if there exist partial words x and y such that $u \subset xy$ and $v \subset yx$ [9].

Proposition 21. *Let $X \subset W\backslash\{\varepsilon\}$. If X is a circular pcode, then X does not contain two distinct conjugate partial words.*

Proof. Suppose that there exist two distinct conjugate partial words u and v in X, and let x, y be partial words such that $u \subset xy$, $v \subset yx$. If $x = \varepsilon$ or $y = \varepsilon$, then $u \uparrow v$, contradicting the fact that X is a pcode. So we may assume that $x \neq \varepsilon$ and $y \neq \varepsilon$. Since X is a circular pcode, the two conditions $yux \uparrow vv$ and $u \uparrow xy$ imply $x = \varepsilon$, a contradiction. ■

Proposition 22. *Let $X \subset W\backslash\{\varepsilon\}$ be a circular pcode. If $u \in X$, then u is primitive.*

Proof. Suppose that there exist $u \in X$ and a partial word v such that $u \subset v^n$ with $n \geq 2$. It follows that $vuv^{n-1} \uparrow uu$ and $u \uparrow v^{n-1}v$. Since X is a circular pcode, then $v^{n-1} = \varepsilon$. We conclude that $v = \varepsilon$, a contradiction. ■

Acknowledgements
This material is based upon work supported by the National Science Foundation under Grant CCR-0207673. I thank the referees of a preliminary version of this paper for their very valuable comments and suggestions.

References:

[1] [1] J. Berstel, L. Boasson, Partial words and a theorem of Fine and Wilf, Theoret. Comput. Sci. 218 (1999) 135–141.
[2] J. Berstel, D. Perrin, Theory of Codes, Academic Press, New York, 1985.
[3] F. Blanchet-Sadri, On unique, multiset, and set decipherability of three-word codes, IEEE Trans. Inform. Theory 47 (2001) 1745–1757.
[4] F. Blanchet-Sadri, A periodicity result of partial words with one hole, Comput. Math. Appl. 46 (2003) 813–820.
[5] F. Blanchet-Sadri, Primitive partial words, preprint 2003.
[6] F. Blanchet-Sadri, Periodicity on partial words, Comput. Math. Appl. 47 (2004) 71–82.
[7] F. Blanchet-Sadri, A. Chriscoe, Local periods and binary partial words: an algorithm, Theoret. Comput. Sci. 314(2004)189–216.
[8] F. Blanchet-Sadri, R.A. Hegstrom, Partial words and a theorem of Fine and Wilf revisited, Theoret. Comput. Sci. 270 (2002) 401–419.
[9] F. Blanchet-Sadri, D.K. Luhmann, Conjugacy on partial words, Theoret. Comput. Sci. 289 (2002) 297–312.
[10] C. Choffrut, J. Karhumäki, Combinatorics of words, in: G. Rozenberg, A. Salomaa (Eds.), Handbook of Formal Languages, Vol. 1, Springer, Berlin, 1997, pp. 329–438 (Chapter 6).
[11] F. Guzmán, Decipherability of codes, J. Pure Appl. Algebra 141 (1999) 13–35.
[12] L.H. Haines, On free monoids partially ordered by embedding, J. Combin. Theory 6 (1969) 94–98.
[13] T. Head, A. Weber, Deciding multiset decipherability, IEEE Trans. Inform. Theory 41 (1995) 291–297.
[14] G. Higman, Ordering with divisibility in abstract algebras, Proc. London Math. Soc. 3 (1952) 326–336.
[15] M. Ito, H. Jürgensen, H.J. Shyr, G. Thierrin, Anti-commutative languages and *n*-codes, Discrete Appl. Math. 24(1989)187–196.
[16] M. Lothaire, Combinatorics on Words, Addison-Wesley, Reading, MA, 1983.

[17] J. Setubal, J. Meidanis, Introductionto Computational Molecular Biology, PWS Publishing Company, Boston, MA, 1997.

[18] H.J. Shyr, Free Monoids and Languages, Hon Min Book Company, Taichung, Taiwan, 1991.

[19] H.J. Shyr, G. Thierrin, Codes and binary relations, Séminaire d'Algèbre, Paul Dubreil, Paris, 1975–1976, Lecture Notes of Springer, Vol. 586, Springer, Berlin, pp. 180–188.